Author's response to reviews

Title: Psychometric evaluation of the community reintegration of injured service members computer adaptive test (CRIS-CAT)

Authors:

Linda Resnik (Linda_Resnik@Brown.edu)
Matthew Borgia (Matthew.Borgia@va.gov)
Pengsheng Ni (pсли@bu.edu)
Paul Pirraglia (Paul.Pirralgia@va.gov)
Alan Jette (Ajette@bu.edu)

Version: 3 Date: 16 November 2011

Author's response to reviews: see over
EDITORIAL REQUESTS:

Experimental research that is reported in the manuscript must have been performed with the approval of an appropriate ethics committee. Research carried out on humans must be in compliance with the Helsinki Declaration (http://www.wma.net/e/policy/b3.htm), and any experimental research on animals must follow internationally recognized guidelines. A statement to this effect must appear in the Methods section of the manuscript, including the name of the body which gave approval, with a reference number where appropriate.

The research was fully approved by our IRB. We added a statement to this effect to the methods section.

Informed consent to participate in the study must also be documented. Manuscripts may be rejected if the editorial office considers that the research has not been carried out within an ethical framework, e.g. if the severity of the experimental procedure is not justified by the value of the knowledge gained.

Informed consent was obtained from all study participants. We added a statement to this effect to the methods section.

REVIEWER 1
Reviewer: Pim Kuipers

Respondent burden questions were overwhelmed by data and psychometric properties. The article did not seem to deal with issues of respondent burden of CRIS, respondent burden of a CAT, etc

We added an additional sentence on P. 6 of the introduction highlighting the benefits of CATs in reducing burden. In this paper we assess respondent burden of CRIS-CAT by comparing the number of items administered in our CAT (administration study) to the number of items in the longer CRIS fixed form.

Methods are appropriate for psychometric properties article, but in my opinion suffer from presentation of too many findings, and not enough sifting and discussion of relevant aspects

We appreciate the reviewer’s comment and have attempted to simplify the organization of the paper to make it easier to follow. We have revised the discussion to highlight the most relevant points and discuss the findings in the context of prior research.

I did not note ethics committee approval information - presumably this is in a related article

We added this information to the methods section.

Are the discussion and conclusions well balanced and adequately supported by the data? - yes somewhat supported by the data, but could have more substance (For example, the hypotheses related to employment, housing, and QoL changes after one year. The discussion about 1 year changes barely mentions
employment, doesn't explicitly mention housing and QoL, and introduces marital status).
We agree that the Discussion section would benefit from expansion and have augmented the discussion on the small number of persons in our sample who had changes in employment and marital status. We also discuss changes in housing stability and QoL. Each of these variables was hypothesized to be associated with community reintegration scores, as explained on p. 11.

Likewise, the authors attribute the lack of relationship to short time frame and low numbers, but these seem quite substantial in comparison to other studies. This is an important finding, with considerable relevance to the substance of the measures being evaluated requires, and I think requires more attention. We provided data comparing the divorce rate in our cohort to rates published in the literature. We have added to the discussion by stating that our results are not likely generalizable to all OEF/OIF Veterans.

Do the title and abstract accurately convey what has been found? yes (though in my opinion requires more discussion / interpretation and less data)
We are open to suggestions regarding changes to the study title. We believe that the abstract accurately reflects the methods and results of the study.

Is the writing acceptable? mostly - in my copy the figure numbers did not correspond.
We checked our figure numbers against the text and all appear correct.

I found the article very complicated - field study, cohort study, administration study, cohort A, cohort B, cohort C.
We reorganized the methods and results sections so that the sub-studies were described separately. We think that this make the paper easier to follow. We also added a new Figure that shows the relationship of the 3 sub-studies (Figure 1)

Some confusing sentences:
P2 abstract - methods first sentence is confusing
We rewrote this sentence as follows: “This was a three-part study that included a 1) a cross-sectional field study of 517 homeless, employed, and Operation Enduring Freedom / Operation Iraqi Freedom(OEF/OIF) veterans ;who completed all items in the CRIS item set, 2) a longitudinal cohort study with one year follow-up study of 135 OEF/OIF veterans, and 3) a 50-person study of CRIS-CAT administration.”

P5 ICF title incorrect
The ICF title was corrected to read, “International Classification of Functioning, Disability and Health”

P7 second last sentence – edit
P7 (and throughout) "data" is plural - should be "data were used"
We corrected this error.
The title suggests a psychometric evaluation of the CRIS-CAT. Because it is stated that "In previous work we developed the CRIS CAT...", and since there is no reference to this I would expect that this paper was in fact THE psychometric evaluation. But it is not. …

The external validity of the CRIS-CAT can only be evaluated after it has been shown that the meaning of its scales are unambiguous, i.e. are unidimensional. This manuscript is a companion to another paper that is now In Press at the Journal of Rehabilitation Research and Development (JRRD). That paper explains the validation of the IRT models and provides evidence of the unidimensionality of the scales. We have made this clearer in the introduction and have provided a citation to the In-Press paper, which we hope that we can update with the actual citation if it is published, before this manuscript is finalized.

The three scales of CRIS-CAT, of which no examples were provided, seem to consist of items which, in my opinion, are causal indicators instead of effect indicators (Fayers and Hand, 1997). So maybe CAT is not appropriate at all. The authors may be much more interested in short assessment not assuming latent constructs such as, for example, Finkelman et al. (2011)

We included an Appendix that contains the CAT items for each of the 3 scales. Similar to the items in the SF-36 (as an example) we believe that CRIS-CAT items are effect (reflective) indicators and as such are appropriate for modeling through short assessment.

I think the paper has structure which makes it very difficult to read. Often it is not explained why samples, external measures etc are used. For example, it is not really clear why the CRIS-CAT needs to be able to differentiate between the three groups. In addition, it is not made clear why a study of the minimum detectable change is necessary; I can make up these things myself after thinking for a while, but it is the work of the authors to do that.

As explained above, we have reorganized the paper to make it easier to follow. We have also explained the rationale for distinguishing the three groups (known groups validation) (p. 8), and calculating the minimal detectable change (p 16).
Some statistical/measurement concepts are described improperly. On page 13, the authors talk about true variance; variance of what? a true score, or a latent trait? They go on to introduce an equation which has the word "standard error" in it. A standard error of what parameter, using which model? Same holds for equations on pages 14-15.

We edited the equation to read

\[
SEM = [SD \text{ between subject scores at 1st Assessment}] \times [ (1-[r \text{ (i.e. ICC)}])]^{0.5} \\
MDC = [z \text{ score (for 90\% confidence)}] \times [SEM]^{2^{0.5}}
\]

Likewise, on pages 6-7 (and figures 1-2), reliability appears in a CAT discussion. What type of reliability do you mean, marginal reliability, or internal consistency reliability such as Cronbach’s? IRT uses information, which is a conditional estimate of measurement precision.

In this comment, Figure 2 is conditional reliability \( Reliability=1-SE^2 \). We have added a reference to this type of reliability.

**Minor Essential Revisions**

References [37], on page 13, and [30] on page 19 are not in the reference list.
These references were added to the reference list.

There is a 30\% dropout rate for the longitudinal study. Please state why missings need not be dealt with. Alternatively, use ML estimation or an alternative to deal with this.
We are not aware of a method to deal with missing outcome data in a small longitudinal cohort study.

**REVIEWER 3**
**Reviewer:** Morten Aa. Petersen

**Major Compulsory Revisions**

Methods, Data Collection, Study Measures: “At Visit 1 we collected data on other measures which we believed would be correlated with CRIS-CAT scores if the instruments were measuring community integration...” Given that almost anything correlates and that the set of measures used for these correlations are not measures of community integration, this seems as a rather vague hypothesis and as rather vague evidence that the CRIS-CAT measures community integration. For example, PCS and MCS may be significantly correlated in many samples, but this seems as poor evidence that PCS measures physical performance and MCS mental. Ideally, one would use some kind of “gold standard” for such correlations, however, if such is not available (which is rarely the case), then I think the authors at least should set up some more specific hypothesis of what would be expected of a measure of community integration.
For example, how do they expect the CRIS-CAT measures to correlate with each of the other measures (high, moderate, or low correlation), do they expect the
three CRIS-CAT scales to correlate similarly with the other measures? Otherwise, I think the conclusions regarding (concurrent) validity should be toned down. We expected to find levels of correlation similar to what we identified in our earlier studies of the CRIS fixed form measure (explained in a new section on p. 14). We added some discussion of our findings in a new paragraph found on p.20.

**Minor Essential Revisions**

Methods, 3rd paragraph, line 4: “In previous work we developed the CRIS CAT by…” No references are given to this previous work. If this work is published please provide references.

We have provided a citation to the manuscript which has just been accepted for publication in JRRD. This paper provides full information on the development of the CRIS-CAT.

Results, Cohort Study Sample & Table 2: There were several findings of significant differences between the participants in the cohort study and those lost to follow-up. Hence, the participants seem as a selected, non-representative sample of OEF/OIF veterans. How does this affect the generalisability of the findings?

All subjects in the cohort study were OEF/OIF Veterans. Those subjects lost to follow up were younger, more recently deployed, more likely to be unemployed, have lower incomes, and be unmarried. They were less likely to have children, depression, alcohol/drug abuse or new mental health diagnosis. We agree that these differences could diminish the generalisability of our findings to those OEF/OIF veterans who are younger and more recently returned from deployment. We discuss the sample differences due to loss to follow-up as a study limitation.

Figure 4: What is meant by RRR on the x-axis?

We changed RRR to “relative risk ratios” in the x axis and OR to “odds ratio” to diminish possible confusion.

Discussion, 5th paragraph, line 2: “…can be administered with minimal respondent burden.” Given the quite limited evaluations of response burden (only the number of items required for one CAT setting is reported, no response time, patient feedback etc. are reported), this seems as quite a strong statement. This needs further justification or should be toned down.

This sentence was edited as follows, “Our study of CAT administration demonstrated that the three scales have excellent test-retest reliability and were administered with reduced respondent burden (as gauged by number of items) as compared to the full item set and the CRIS fixed form measure.”

**Discretionary Revisions**

Methods, Statistical Analyses, 4th paragraph, line 6: “We ran 3 separate multinomial regression models to predict change in…” Why three separate regression and not one multiple regression including all three CRIS-CAT scales? This would elucidate which CRIS-CAT scale might be the strongest predictor.
Thank-you for this comment. We believe that there is value in modeling the outcomes using each CRIS-CAT scale alone, as we anticipate that the CRIS-CAT subscales will be used individually and in combination (depending on the clinical or research needs). That said, we did construct multivariable regression models including all three scales, per your recommendation. We found that the magnitude of the beta coefficients were not changed but their significance was lost. This, we believe, was due to the high correlation between the subscales. For these two reasons we would rather report the results of the separate regression models.

Table 2: It would be more informative if the actual p-values, and not just *'were shown (this is done in table 3), possibly highlighting significant findings, e.g. p-values in bold.
Thank you for this suggestion. We included the actual p-values and presented significant p-values in bold to help highlight the findings of our study.

Table 5: Assuming that the SF12 v1 measure used is PCS v1 when predicting PCS v2 and MCS v1 for MCS v2, it is surprising to me that in all but one case, the CRIS-CAT scales at visit 1 seem as better predictors of PCS/MCS v2 than PCS/MCS v1. Was this really the case? If so, please try to explain/comment on this.
Although the beta coefficients are slightly higher, the confidence intervals of these estimates overlap and thus we cannot say that the CRIS-CAT was more predictive of visit 2 SF12 scores than were visit 1 SF12 scores. It appears that they are equivalent predictors.

That said, we realized that this table could be improved based on your comment and so we reorganized it to show the coefficients for the MCS and PCS at visit 1 as predictors of the MCS and PCS scores at follow-up.

Minor issues not for publication
Methods, line 2: “of OEF/OIF veterans” is written twice
This was corrected

Methods, line 7: “examination the CRIS-CAT’s”, missing “of” after examination
This was corrected

Statistical Analyses, 2nd paragraph, last line: “1/(1+(standard error)2 )[37]” seems to refer to a reference 37, but there are only 22 references in the list of references. This reference was added and the numbering of the references was corrected.

Discussion, 6th paragraph, line 6: “overestimate these correlations. [30]” seems to refer to a reference 30, but there are only 22 references in the list of references. This reference was added and the numbering of the references was corrected.

Figure 2a. “Reliability of CAT estimates: Extent Scale” is labeled 1a.
The numbering of the Figure was corrected (these are now Figures 3a-c) and the titles amended to indicate that these figures show conditional reliability.

**Figure 2a-c: Cannot read the titles for the y-axes.**
We changed the font of the axes (now numbered Figures 3a-c), hopefully they are now more readable.